PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Suitability of emergency department attenders to be assessed in primary care: survey of general practitioner agreement in a random sample of triage records analysed in a service evaluation project.
AUTHORS	Lasserson, Daniel; Thompson, Mary; McCann, Lloyd; Thompson, Matthew; Heneghan, Carl

VERSION 1 - REVIEW

REVIEWER	Thomas E. Cowling Department of Primary Care and Public Health Imperial College London U.K.
REVIEW RETURNED	18-Aug-2013

THE STUDY	Abstract Page 2, Line 13: The statement does not indicate the design of the study; it describes the study population. Article Summary Page 3, Line 23: According to the 'Abstract' and 'Results' sections, the mean percentage of attendances deemed appropriate for primary care management was 43%, not 40% as written. Introduction Page 4, Line 6: 'The last decade has seen a substantial increase in the number of patients attending hospital emergency departments (EDs)' requires indication of the setting of these EDs. In the article of mine cited, I refer only to hospital EDs in England and the U.S. In the case of England, I report the trend from 2007-08 to 2011-12 (not the last decade), and the U.S. study I cite reports the trend from 1997 to 2007 (also not the last decade). Therefore, the accuracy of the opening statement could be improved. It may also be preferable to cite the original sources, the Health and Social Care Information Centre and Tang et al., rather than my article. Page 4, Line 8: '…consequently the delivery of emergency healthcare in England is under significant threat' attributes the 'significant threat' to rising demand. However, demand is only one parameter in a complex system; it is perhaps too simplistic to make the attribution written. Page 4, Line 25: 'The association between reduced access to primary care and increases in ED attendance in cross-sectional data…,' where an article of mine is cited, could be more accurately re-written as 'The cross-sectional association between less timely

	access to primary care and greater rates of self-referred discharged ED attendances….' Page 4, Line 31: Where 'there is evidence that demand can be reduced by increasing access to primary care outside normal office hours' is written, it may be preferable to indicate, at least, the design and setting of the study cited, as the study was conducted in one region of the Netherlands in 2001/2 and employed a before-and-after design. Page 4, Line 35: An additional citation for 'reasons for attendance at an ED with non-urgent problems' is Agarwal S, Banerjee J, Baker R, Conroy S, Hsu R, Rashid A, et al. Potentially avoidable emergency department attendance: interview study of patients' reasons for attendance. Emerg Med J 2012;29:e3. doi: 10.1136/emermed-2011-200585. Page 4, Line 37: '…a number of initiatives have been trialled within EDs once patients with non-urgent presentations attend…' could be more clearly phrased as '…a number of interventions designed for patients who present with non-urgent problems have been trialled within EDs….' Page 4, Line 40: The abbreviation 'GP' is used for the first time in the main text and could be introduced as '…primary care physicians (general practitioners; GPs)' as both 'primary care physicians' and 'GPs' are used subsequently in the manuscript. Page 4, Line 46: '… particularly as non-urgent presentations receive less investigations and follow-up if managed by a GP' is written too definitively. The existing evidence on this subject is inconsistent and inconclusive. Page 5, Line 8: One aim is written as 'to measure the level of agreement among primary care physicians about which types of patients who attend ED[s] could be appropriately managed in primary care….' This aim is not addressed in the study; however, the study does measure the level of agreement among primary care physicians about who within different types (adult/paediatric; trauma/medicine/surgery) of patient could be appropriately managed in primary care. The unit of GP assessment is not the type, it is the patient. Methods Page 5, Line 18: It may be useful to provide some additional information on what the 'data extraction tool' was and how it functioned. Further, how was the 'reason for presentation' identified and how many 'reasons' for presentation were considered? Page 5, Line 23: How did the primary care physicians demonstrate that the information extracted was sufficient to reach decisions? Page 5, Line 31: It may be preferable to describe the setting before detailing the methods of the pilot on 20 cases. Also, it could be useful to provide some information about the design of the ED and surrounding health services. Is it solely a type 1 A&E department, a consultant-led 24 hour service with full resuscitation facilities, or does it, for example, have a co-located type 3 A&E department,

	such as an urgent care centre? Are there minor injuries units or walk-in centres close to John Radcliffe Hospital that may affect demand? Page 5, Line 39: Did the questionnaires also contain information on demographics, reason for presentation, and triage nurse assessment (the information extracted in the pilot and deemed sufficient to reach decisions on suitability)? Page 6, Line 8: It is written that 'Agreement results are presented as proportions and ranges.' It would appear that they are presented as kappa statistics and 95% confidence intervals however, as indicated in the top-left cell of Table 1. Page 6, Line 8: The justification for dichotomising the responses would be useful to provide. If there were three categories of response, rather than two, I would expect the kappa statistics to be smaller than the ones reported. Therefore, the measure would indicate lower inter-rater reliability. As the reader, I would be interested to know how much the dichotomisation affects the results. References Page 10, Line 53: What is currently written as 'England N.', I presume should be 'NHS England'. General At several points in the manuscript, 'ED' is written when 'the ED' (referring to the emergency department at John Radcliffe Hospital) or 'EDs' (referring to emergency departments in the U.K.) may be preferable. For example, on page 2, lines 41 and 50, 'patients attending ED could' and 'pressures in ED in the U.K.' are written. In the Methods section, the authors have written that 'Each case was assessed independently by each physician for 1) appropriateness for primary care management….' However, at other points in the manuscript, 'appropriateness for treatment' (for example, page 2, line 23) and 'can be seen in primary care' (page 3, line 18) are referred to in the same context. Management, treatment, and 'being seen' may not be the same thing, so perhaps the consistent use of terms is advisable.
RESULTS & CONCLUSIONS	Results Page 6, Line 31: It is stated that the GPs made a decision on suitable location of treatment in 1,291 responses. However, only 629 cases were allocated to the GPs for assessment, with each case reviewed twice (by both members of a GP Pair). Since, 2*629=1,258, how were 1,291 responses provided? Page 6, Line 35: According to the abstract, 43% was the mean percentage of patients deemed suitable for primary care management; it may be helpful if the first sentence of the second paragraph in the Results section explicated that the 43% figure refers to the mean. This point also applies to the sentence on page 7, line 21. Page 6, Line 40: The addition of '95% CI' within the parentheses is required.

	Page 7, Line 5: Where '…significantly fewer patients required investigations…' is written, the wording should reflect that this is the retrospective assessment of the GPs and not whether the patients did actually, for example, require investigations. Page 7, Lines 10-16: The use of 'versus' or 'vs.' should be consistent. Page 7, Table 2: The presentation of the first column is inconsistent with the presentation of the first column in Table 1 on page 6. Discussion Page 8, Line 31: It would be useful to provide the 95% confidence interval for the kappa statistic provided. Page 9, Line 37: A further limitation could be that the sample consisted only of patients who attended the ED in one month, November 2008. The results may not be generalisable to patients attending the ED in, for example, other months or across the whole year. Page 9, Line 46: Typographical error: '…is less then the 40%…' is written. Also, the mean percentage estimated in the study is 43%, so I wonder why 40% is given.
REPORTING & ETHICS	Page 10, Line 21: It states that a grant was provided to 'four primary care physicians to conduct the audit.' Are these the same four GPs that assessed patients' suitability for management in a primary care setting in the study? Page 10, Line 33: Why was ethical approval not required, given that patients' case notes were retrieved?

REVIEWER	Jane E.Bickerton Lecturer Applied Biological Diagnostic and Therapeutic Sciences School of Health Sciences, 1 Myddelton Street London, EC1R 1UW.
REVIEW RETURNED	20-Aug-2013

THE STUDY	The aim of this study is to assess the level of agreement between four GPs working in an emergency department (ED) evaluating which attendees could be treated in the ED, a primary care setting, or specialist review. The GP decision is based on information about the presenting complaint, demographics and an ED triage nurse assessment. The study outcomes were useful. The article strengthens the need for clear definitions of not only 'triage' but also terms such as 'signposting' and 'streaming'. The choice of specialist review, ED and primary care services is limited. Given the breadth of Primary health and social care today it would have made sense to differentiate between primary care services, non- national health services as well as ED. Reference Bickerton, J , Davies, J, Davies, H, Apau, D, and Procter, S. (2011).

	Streaming primary urgent care: a prospective approach. Primary Health Care Research & Development 13: 142–152 doi:10.1017/S146342361100017X
RESULTS & CONCLUSIONS	Bickerton et al (2012) reporting on streaming in ED and adjacent Walk in Centres (WiC) or a Primary Care Urgent Care Centre (PUCC) found a 30.5% level of agreement between a GP, a nurse consultant and a community pharmacist which was lower than the present study. However, the validation exercise found that professional choices for service outcome were not based on clinical assessment alone, but included personal knowledge of types and professional competence of local services as well as equipment. The decision as to which primary care service was most appropriate was influenced by the convenience of services for the consumer.
REPORTING & ETHICS	The names of the particular research site should be removed
GENERAL COMMENTS	I appreciate the earlier reviews on this article and have added a few minor additions and thank the authors for their enlightening article.

REVIEWER	Dr Caroline Anne Mitchell MBChB MD FRCGP PGCertMEd GP and Senior Clinical Lecturer Academic Unit of Primary Medical Care University of Sheffield Sheffield United Kingdom
REVIEW RETURNED	22-Aug-2013

THE STUDY	There is no supplemental information which would raises questions about the work, nor change th emanuscript
RESULTS & CONCLUSIONS	
REPORTING & ETHICS	
GENERAL COMMENTS	A timely observational study which provides useful research data to support the devolopment and evaluation of alternative models of healthcare delivery for unscheduled NHS care. This article will be of interest to primary care clinicans. emergency department specialists and comiisioners of healthcare.

REVIEWER	Nicholas Steel Clinical Senior Lecturer in Primary Care Norwich Medical School UK
REVIEW RETURNED	23-Aug-2013

THE STUDY	This study reports on the proportion of 629 patients who had attended Oxford emergency department that were considered suitable for primary care, based on a retrospective review of clinical notes by 4 GPs. This is an important topic, and the authors refer to published results on the subject, including tools to describe appropriateness of care from clinical records. They conclude that generating consensus on methods to identify patients who could be managed in primary care is a major priority. Given this, it seems a missed opportunity to simply describe the results of this study, without comparing the methods and results to existing tools. The important question to be answered from this research is whether this GP assessment is an improvement on

	existing methods for identification of emergency dept patients who could be safely managed in primary care. The paper as written does not address this, but I can't see any reason why it could not be revised to do so. More information is needed on the 'pilot data extraction tool' - how was it developed, what was it, and what information did it extract? How did the two primary care physicians 'demonstrate that this information was sufficient...'? (first paragraph of methods section). Why did the authors choose to develop a new tool rather than use the existing tool they reference in the discussion, and how does the finished version compare? It would be helpful if the questionnaire referred to could be included as an appendix. Who were the 4 GPs (full-time, academic, new salaried GPs, older partners, etc?) How did they make their assessment of 'appropriateness for primary care management'? Were they given any guidance or criteria? I did not understand why the sample size estimation was done on kappa 0.6 vs kappa 0.8? The abstract concludes that 'stronger agreement may be seen for paediatric than for adult attenders', and that 'there is now urgent need to implement more effective signposting of patients', yet neither of these is fully supported by the data presented. The stronger agreement has fairly wide confidence intervals that overlap between the 2 categories, and the study did not report any research into 'signposting'.
RESULTS & CONCLUSIONS	The 2 tables present only kappas, and the main outcome is only given in figure 2. It would be helpful to have a table describing the characteristics of the sample, and a table presenting the main results (ie proportion deemed suitable for primary care).
REPORTING & ETHICS	The authors state that ethical approval was not required. I agree in principle, but it would be helpful to have a few more details to support this somewhere in the paper. For example, how were the data extracted and at what stage were they anonymised? Was the triage nurse mentioned in the methods section involved in this data extraction?

VERSION 1 – AUTHOR RESPONSE

Reviewer: Thomas E. Cowling
Department of Primary Care and Public Health Imperial College London U.K.

1. Page 2, Line 13: The statement does not indicate the design of the study; it describes the study population.

We have altered the statement on study design to include the description of the survey.

2. Page 3, Line 23: According to the 'Abstract' and 'Results' sections, the mean percentage of attendances deemed appropriate for primary care management was 43%, not 40% as written.

We have altered this to 43%.

3. Page 4, Line 6: 'The last decade has seen a substantial increase in the number of patients attending hospital emergency departments (EDs)' requires indication of the setting of these EDs. In the article of mine cited, I refer only to hospital EDs in England and the U.S. In the case of England, I report the trend from 2007-08 to 2011-12 (not the last decade), and the U.S. study I cite reports the trend from 1997 to 2007 (also not the last decade). Therefore, the accuracy of the opening statement could be improved. It may also be preferable to cite the original sources, the Health and Social Care Information Centre and Tang et al., rather than my article.

We have altered the sentence to read " There has been a substantial increase in the number of patients attending hospital emergency departments over the last six years in England" We have kept the reference to Cowling's article as it is a good introduction to the area for the general reader.

4. Page 4, Line 8: '…consequently the delivery of emergency healthcare in England is under significant threat' attributes the 'significant threat' to rising demand. However, demand is only one parameter in a complex system; it is perhaps too simplistic to make the attribution written.

We thank this reviewer for this point. We have changed this sentence to "The delivery of emergency healthcare in England is under significant threat currently, partly due to rising demand, and improvements to emergency care provision is now a major NHS priority".

5. Page 4, Line 25: 'The association between reduced access to primary care and increases in ED attendance in cross-sectional data…,' where an article of mine is cited, could be more accurately re-written as 'The cross-sectional association between less timely access to primary care and greater rates of self-referred discharged ED attendances….'

We have altered the sentence accordingly.

6. Page 4, Line 31: Where 'there is evidence that demand can be reduced by increasing access to primary care outside normal office hours' is written, it may be preferable to indicate, at least, the design and setting of the study cited, as the study was conducted in one region of the Netherlands in 2001/2 and employed a before-and-after design.

We have altered the sentence accordingly.

7. Page 4, Line 35: An additional citation for 'reasons for attendance at an ED with non-urgent

problems' is Agarwal S, Banerjee J, Baker R, Conroy S, Hsu R, Rashid A, et al. Potentially avoidable emergency department attendance: interview study of patients' reasons for attendance. Emerg Med J 2012;29:e3. doi: 10.1136/emermed-2011-200585.

We thank the reviewer for this citation which we have inserted at this point in the text.

8. Page 4, Line 37: '…a number of initiatives have been trialled within EDs once patients with non-urgent presentations attend…' could be more clearly phrased as '…a number of interventions designed for patients who present with non-urgent problems have been trialled within EDs….'

We have altered the sentence accordingly.

9. Page 4, Line 40: The abbreviation 'GP' is used for the first time in the main text and could be introduced as '…primary care physicians (general practitioners; GPs)' as both 'primary care physicians' and 'GPs' are used subsequently in the manuscript.

We have altered the sentence accordingly.

10. Page 4, Line 46: '… particularly as non-urgent presentations receive less investigations and follow-up if managed by a GP' is written too definitively. The existing evidence on this subject is inconsistent and inconclusive.

We have altered this sentence so that it is less definitive and it now reads "particularly as some studies suggest that non-urgent presentations receive less investigations and follow up if managed by a GP"

11. Page 5, Line 8: One aim is written as 'to measure the level of agreement among primary care physicians about which types of patients who attend ED[s] could be appropriately managed in primary care….' This aim is not addressed in the study; however, the study does measure the level of agreement among primary care physicians about who within different types (adult/paediatric; trauma/medicine/surgery) of patient could be appropriately managed in primary care. The unit of GP assessment is not the type, it is the patient.

We thank this reviewer for this clarification. We have altered the aim to read "Our aim in this study was to measure the level of agreement among primary care physicians about who could be appropriately managed in primary, within different groups of patients (based on age range or clinical categories of trauma, medicine or surgery) care and to assess whether agreement differed between these groups."

12. Page 5, Line 18: It may be useful to provide some additional information on what the 'data extraction tool' was and how it functioned. Further, how was the 'reason for presentation' identified and how many 'reasons' for presentation were considered?

In the pilot, we utilised a word document template as the data extraction tool, based on that used in Lowy A, Kohler B, Nicholl J Attendance at accident and emergency departments: unnecessary or inappropriate? J Public Health Med. 1994 16(2):134-400). This template included 'Presenting complaint' extracted from the Emergency Department Casualty Card / ED Notes and only the primary presenting complaint was included. We have now included the reference to this tool and thank the reviewer for requesting clarification of this point. The sentence now reads "We developed a pilot data extraction tool based on that used by Lowy et al25 using information from the initial ED presentation of adult and child patients, including demographics, reason for presentation and triage nurse assessment on an initial 20 cases." The reasons for presentation are those written in the ED notes,

i.e. they were not limited to a priori categories, as implied by the sentence.

13. Page 5, Line 23: How did the primary care physicians demonstrate that the information extracted was sufficient to reach decisions?

The primary care physicians made their decisions based on the information available, as stated in the methods.

14. Page 5, Line 31: It may be preferable to describe the setting before detailing the methods of the pilot on 20 cases. Also, it could be useful to provide some information about the design of the ED and surrounding health services. Is it solely a type 1 A&E department, a consultant-led 24 hour service with full resuscitation facilities, or does it, for example, have a co-located type 3 A&E department, such as an urgent care centre? Are there minor injuries units or walk-in centres close to John Radcliffe Hospital that may affect demand?

We have inserted the sentence “The John Radcliffe Hospital emergency department is a consultant-led, 24 hour service with full resuscitation facilities without a co-located urgent care centre or nearby walk-in centre.” to address this point.

15. Page 5, Line 39: Did the questionnaires also contain information on demographics, reason for presentation, and triage nurse assessment (the information extracted in the pilot and deemed sufficient to reach decisions on suitability)?

The questionnaires contained the information that was supplied in the pilot data extraction as described above. We have amended the sentence to “We extracted data established as satisfactory in the pilot for GP decision making from the record documented by a triage nurse and transferred this to an electronic questionnaire”.

16. Page 6, Line 8: It is written that ‘Agreement results are presented as proportions and ranges.’ It would appear that they are presented as kappa statistics and 95% confidence intervals however, as indicated in the top-left cell of Table 1.

Raw agreement results are presented as proportions and ranges in the text, and kappa statistics with 95% CI are presented in the tables. We have altered this sentence to clarify this particular issue.

17. Page 6, Line 8: The justification for dichotomising the responses would be useful to provide. If there were three categories of response, rather than two, I would expect the kappa statistics to be smaller than the ones reported. Therefore, the measure would indicate lower inter-rater reliability. As the reader, I would be interested to know how much the dichotomisation affects the results.

Dichotomising response as ‘yes’ vs. ‘no’ and ‘unsure’ conservatively estimates the raw proportion who are suitable for primary care management, an important outcome of this study. We have altered this sentence to clarify this point. An additional sensitivity analysis based on a post hoc alteration of our methods would not significantly add to the message of the paper and it is entirely predictable that increasing levels of response reduces kappa values.

18. Page 10, Line 53: What is currently written as ‘England N.’, I presume should be ‘NHS England’.

We thank this reviewer for noticing this error, which we have now corrected.

19. At several points in the manuscript, ‘ED’ is written when ‘the ED’ (referring to the emergency

department at John Radcliffe Hospital) or ‘EDs’ (referring to emergency departments in the U.K.) may be preferable. For example, on page 2, lines 41 and 50, ‘patients attending ED could’ and ‘pressures in ED in the U.K.’ are written.

We have corrected these examples as this reviewer suggests.

20. In the Methods section, the authors have written that ‘Each case was assessed independently by each physician for 1) appropriateness for primary care management….’ However, at other points in the manuscript, ‘appropriateness for treatment’ (for example, page 2, line 23) and ‘can be seen in primary care’ (page 3, line 18) are referred to in the same context. Management, treatment, and ‘being seen’ may not be the same thing, so perhaps the consistent use of terms is advisable.

We thank this reviewer for this point, and ‘management’ is now used consistently.

21. Page 6, Line 31: It is stated that the GPs made a decision on suitable location of treatment in 1,291 responses. However, only 629 cases were allocated to the GPs for assessment, with each case reviewed twice (by both members of a GP Pair). Since, 2*629=1,258, how were 1,291 responses provided?

We have corrected this error.

22. Page 6, Line 35: According to the abstract, 43% was the mean percentage of patients deemed suitable for primary care management; it may be helpful if the first sentence of the second paragraph in the Results section explicated that the 43% figure refers to the mean. This point also applies to the sentence on page 7, line 21.

We have altered these sentences accordingly “The mean GP assessment of ED attendances suitable for primary care management was 43% (range 38% to 47%).” and “The mean GP assessment on suitability for primary care management was 42% in adults (range 36% to 49%) and 48% in children (range 40% to 57%).”

23. Page 6, Line 40: The addition of ‘95% CI’ within the parentheses is required.

This has been added.

24. Page 7, Line 5: Where ‘…significantly fewer patients required investigations…’ is written, the wording should reflect that this is the retrospective assessment of the GPs and not whether the patients did actually, for example, require investigations.

We have changed this to “…significantly fewer patients were deemed to require investigations…”

25. Page 7, Lines 10-16: The use of ‘versus’ or ‘vs.’ should be consistent.

We have now used ‘vs.’ consistently.

26. Page 7, Table 2: The presentation of the first column is inconsistent with the presentation of the first column in Table 1 on page 6.

The first columns are now consistent

27. Page 8, Line 31: It would be useful to provide the 95% confidence interval for the kappa statistic provided

This result is reported in Gribben B. General practitioners' assessments of the primary care caseload in Middlemore Hospital Emergency Department. N Z Med J 2003; 116(1169): U329. The confidence interval is not reported in the full text of the paper.

28. Page 9, Line 37: A further limitation could be that the sample consisted only of patients who attended the ED in one month, November 2008. The results may not be generalisable to patients attending the ED in, for example, other months or across the whole year.

We have added a phrase in the existing sentence in limitations section which now reads “In addition, this study was done in one university hospital during one month of the year which may not generalise to other settings”

29. Page 9, Line 46: Typographical error: ‘…is less then the 40%…’ is written. Also, the mean percentage estimated in the study is 43%, so I wonder why 40% is given.

These have been corrected.

30. Page 10, Line 21: It states that a grant was provided to ‘four primary care physicians to conduct the audit.’ Are these the same four GPs that assessed patients’ suitability for management in a primary care setting in the study?

This is correct.

31. Page 10, Line 33: Why was ethical approval not required, given that patients’ case notes were retrieved?

We have inserted the following text into the methods section to clarify this important point, and also “This study was conducted as a service evaluation and data were obtained primarily for audit purposes according to the guidance from the Oxford ORH Trust audit policies. In accordance with the guidance for research in place at the time the study was conducted, research ethics approval was not required for service evaluations such as this study. All personnel involved in handling data were employees of the hospital trust or (then) primary care trust. Data were anonymised and treated according to the standard operating procedures for patient data in place at the Trust and the University of Oxford Department of Primary Care.” Furthermore we have altered the title so that the general reader is fully aware that data were collected as part of a service evaluation project.

Reviewer: Jane E.Bickerton
Lecturer Applied Biological Diagnostic and Therapeutic Sciences School of Health Sciences,
1 Myddelton Street
London, EC1R 1UW.

1. The aim of this study is to assess the level of agreement between four GPs working in an emergency department (ED) evaluating which attendees could be treated in the ED, a primary care setting, or specialist review. The GP decision is based on information about the presenting complaint, demographics and an ED triage nurse assessment.The study outcomes were useful. The article strengthens the need for clear definitions of not only 'triage' but also terms such as 'signposting' and 'streaming'.

We thank the reviewer for this comment that the outcomes were useful and agree that clear definitions are needed for ‘triage’ and ‘signposting’.

2. The choice of specialist review, ED and primary care services is limited. Given the breadth of Primary health and social care today it would have made sense to differentiate between primary care services, non- national health services as well as ED.

We thank this reviewer for raising an important issue. We agree that there are multiple routes to seeking healthcare in the community and we have now acknowledged this in the limitation section of the discussion with the sentence “Finally we did not ask the GPs if they considered other community based health professionals e.g. pharmacists as an appropriate choice of healthcare access given the clinical presentations”.

Reference
Bickerton, J , Davies, J, Davies, H, Apau, D, and Procter, S. (2011). Streaming primary urgent care: a prospective approach. Primary Health Care Research & Development 13: 142–152 doi:10.1017/S146342361100017X

3. Bickerton et al (2012) reporting on streaming in ED and adjacent Walk in Centres (WiC) or a Primary Care Urgent Care Centre (PUCC) found a 30.5% level of agreement between a GP, a nurse consultant and a community pharmacist which was lower than the present study. However, the validation exercise found that professional choices for service outcome were not based on clinical assessment alone, but included personal knowledge of types and professional competence of local services as well as equipment. The decision as to which primary care service was most appropriate was influenced by the convenience of services for the consumer.

We thank this reviewer for this very interesting citation which we have now referenced in the discussion section of the manuscript, along with the point about the level of agreement, with the addition of the sentence “Low levels of agreement among different professionals about appropriateness of different services for patients attending EDs and walk-in-centres have also been reported, but may be explained by the fact that different professional groups were used to determine consensus.28”

4. names of the particular research site should be removed

We disagree that the name of the research site should be removed, as this was a service evaluation and audit.

5. I appreciate the earlier reviews on this article and have added a few minor additions and thank the authors for their enlightening article.

We thank this reviewer for this very positive comment.

Jane Bickerton

Reviewer: Dr Caroline Anne Mitchell
MBChB MD FRCGP PGCertMEd
GP and Senior Clinical Lecturer
Academic Unit of Primary Medical Care
University of Sheffield
Sheffield
United Kingdom

1. A timely observational study which provides useful research data to support the devolopment and evaluation of alternative models of healthcare delivery for unscheduled NHS care.This article will be of interest to primary care clinicans. emergency department specialists and comiisioners of healthcare.

We thank this reviewer for this positive comment.

Reviewer: Nicholas Steel
Clinical Senior Lecturer in Primary Care Norwich Medical School UK

1. This study reports on the proportion of 629 patients who had attended Oxford emergency department that were considered suitable for primary care, based on a retrospective review of clinical notes by 4 GPs. This is an important topic, and the authors refer to published results on the subject, including tools to describe appropriateness of care from clinical records. They conclude that generating consensus on methods to identify patients who could be managed in primary care is a major priority.

We agree with this synopsis of our study report.

2. Given this, it seems a missed opportunity to simply describe the results of this study, without comparing the methods and results to existing tools. The important question to be answered from this research is whether this GP assessment is an improvement on existing methods for identification of emergency dept patients who could be safely managed in primary care. The paper as written does not address this, but I can't see any reason why it could not be revised to do so.

This reviewer raises an important point about the implications of the findings as they stand. We agree that we did not address this suggested additional aim of comparison with existing methods in the present study. This would be an important additional study to undertake, including an analysis of the results of different tools, but this would be more appropriate for a separate research report.

3. More information is needed on the 'pilot data extraction tool' - how was it developed, what was it, and what information did it extract?

Please see the response to Reviewer 1, question 12 above.
4. How did the two primary care physicians 'demonstrate that this information was sufficient...'? (first paragraph of methods section).

Please see the response to Reviewer 1, question 13 above.
5. Why did the authors choose to develop a new tool rather than use the existing tool they reference in the discussion, and how does the finished version compare? It would be helpful if the questionnaire referred to could be included as an appendix.

We chose to update the Lowy et al tool to reflect current practice. Therefore we used the phrase “appropriate for management in primary care” in our questions. The Lowy et al paper is now referenced in the manuscript and the questions that were included are already present in the methods section.

6. Who were the 4 GPs (full-time, academic, new salaried GPs, older partners, etc?)

The GPs all had experience of out of hours care and worked at least 50% of their time in general practice. The full-time and part-time distinction is less relevant to modern primary care practice as most ‘full time’ GP partners will work eight sessions. The group included partners and one academic

GP with a significant clinical workload at 50% FTE. The limitation of the small number of GPs is already mentioned in the appropriate section of the discussion.

7. How did they make their assessment of 'appropriateness for primary care management'? Were they given any guidance or criteria?

GPs used their clinical judgement to decide what is appropriate for primary care management.
8. I did not understand why the sample size estimation was done on kappa 0.6 vs kappa 0.8?

The reliability of clinicians' ratings is an important consideration in areas such as diagnosis and the interpretation of examination findings. [1] There are a number of factors that can influence the Kappa result (prevalence, bias, and non-independent ratings) and the coefficient represents the proportion of agreement greater than that expected by chance. The difference between a kappa of 0.6 and 0.8 is equivalent to the difference between moderate and substantial agreement. The important issue is that prior to undertaking a reliability study, a sample size calculation should be performed so that a study has a stated probability of detecting a statistically significant kappa coefficient or of providing a confidence interval of a desired width. [2] Thus we undertook such a calculation prior to starting this study. We have provided the references below.

[1] Sim J, Wright CC. Phys Ther. 2005 Mar;85(3):257-68.
The kappa statistic in reliability studies: use, interpretation, and sample size requirements.
[2] Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation. Stat Med.1992 ;11:1511–1519.

9. The abstract concludes that 'stronger agreement may be seen for paediatric than for adult attenders', and that 'there is now urgent need to implement more effective signposting of patients', yet neither of these is fully supported by the data presented. The stronger agreement has fairly wide confidence intervals that overlap between the 2 categories, and the study did not report any research into 'signposting'.

We disagree with this point. We do not definitively report that stronger agreement is seen for paediatric than for adult attenders, only that our data support the statement that it ‘may’. Furthermore the implication of our findings are that signposting may be one strategy as patients who could be seen in primary care are currently self-presenting to EDs. However we have altered the last sentence of the abstract which addresses this reviewer’s concern to “More effective signposting of patients presenting with acute or urgent problems, and supporting a greater role for primary care in relieving the severe workflow pressures in ED in the UK are potential solutions.”

10. The 2 tables present only kappas, and the main outcome is only given in figure 2. It would be helpful to have a table describing the characteristics of the sample, and a table presenting the main results (ie proportion deemed suitable for primary care).

We have included the main results in the text in a short paragraph, and the more detailed results i.e. the kappas are given in tabular form which is easier for the general reader to assess. Presenting complaints were not categorised to avoid bias in data extraction and therefore a summary table of such complaints would be very long without additional gain in terms of interpretation of data. We have already included brief demographic data about age in the results section so that the general reader understands the split between paediatric and adult patients in the sample.

11. The authors state that ethical approval was not required. I agree in principle, but it would be helpful to have a few more details to support this somewhere in the paper. For example, how were the data extracted and at what stage were they anonymised? Was the triage nurse mentioned in the methods section involved in this data extraction?

Please see our response to Reviewer 1, question 31 above.

VERSION 2 – REVIEW

REVIEWER	Thomas E. Cowling Department of Primary Care and Public Health Imperial College London U.K
REVIEW RETURNED	06-Oct-2013

GENERAL COMMENTS	I thank the authors for their thorough responses to my previous comments; I do hope that they were helpful. I have some additional minor comments that the authors may wish to swiftly address. Page 5, Lines 24-45: The setting is currently described in two separate instances, at the beginning of the first and second paragraphs under 'Methods'. The authors may wish to move some information about the ED from the beginning of the second paragraph to the beginning of the first paragraph. Page 6, Line 56 to Page 7, Line 3: 'From the clinical information presented, the GPs were able to make a decision...' is written, but, as I understand it, the questionnaire also contained demographic information and the triage nurse assessment. Perhaps the sentence should read, 'From the information presented in the electronic questionnaire....' Page 6, Line 56 to Page 7, Line 5: The following is not clear to me: '...the GPs were able to make a decision on suitable location of treatment in 88% of cases (n=1258 responses).' Does this mean that the GPs provided a 'Yes' or 'No' response (rather than 'Unsure') in 88% of responses, which equals 1,107 (0.88*1258) responses? Perhaps the sentence could be more clearly written. General: There are still instances in the text where suitability for 'treatment' rather than suitability for 'management' are referred to (e.g. Page 2, Line 26; Page 2, Line 31). The phrase 'could be seen' is also still used on occasion (e.g. Page 3, Line 22; Page 3, Line 27; Page 8, Line 40). The authors, in their response, indicated a preference for using the term 'management', so they may wish to ensure its consistent usage.

REVIEWER	Nicholas Steel Clinical Senior Lecturer in Primary Care Norwich Medical School UK
REVIEW RETURNED	23-Oct-2013

GENERAL COMMENTS	This paper now reads very well and I recommend acceptance for publication without further review, after the authors have considered the minor points below: Abstract results: the main result is described as 'mean GP rating of appropriateness for primary care treatment was 43%', which I found slightly difficult to interpret, as it implies that the GP rating was a % score rather than a yes/no decision. How about: 'mean percentage of patients that GPs considered suitable for primary care treatment was 43%' instead? The kappa agreement is given as 0.54, which is the result for the first pair of GPs. The result of 0.47 for the second pair should also be given here. Abstract conclusion: the stronger agreement for paeds is discussed here, and so should probably be mentioned in the abstract results as well. Brief discussion of the low agreement between GPs & the implication that accurate assessment is difficult for individual patients would be useful here, space allowing. Methods para 1: '2 GPs demonstrated that this information was sufficient' would be more accurate as: '2 GPs considered that this information was sufficient' Methods para 2: the new sentence starting 'We extracted data established as satisfactory....' took me a few readings to understand, and the authors may wish to clarify. Results para 1: clarify whether 'minors' area of ED' refers to minor injuries or children's area. Results para 2 & 4: see comment on abstract above about phrasing of main result. Discussion main findings para 1: The authors may wish to rephrase the sentence starting: 'differences between primary care...' as I found it difficult to interpret. See comment under abstract conclusion above, that discussion of the low agreement between GPs & the implication that accurate assessment is difficult for individual patients should be added here. It is an important point for the future design of interventions in this area, that assessment of patient suitability is a complex task, and agreement even between experienced GPs is not high.

VERSION 2 – AUTHOR RESPONSE

REVIEWER 1

I thank the authors for their thorough responses to my previous comments; I do hope that they were helpful. I have some additional minor comments that the authors may wish to swiftly address.
We thank this reviewer for these comments.

Page 5, Lines 24-45: The setting is currently described in two separate instances, at the beginning of the first and second paragraphs under 'Methods'. The authors may wish to move some information about the ED from the beginning of the second paragraph to the beginning of the first paragraph.
We have moved this sentence to the first paragraph.

Page 6, Line 56 to Page 7, Line 3: 'From the clinical information presented, the GPs were able to make a decision...' is written, but, as I understand it, the questionnaire also contained demographic information and the triage nurse assessment. Perhaps the sentence should read, 'From the information presented in the electronic questionnaire....'

We have altered this sentence accordingly.

Page 6, Line 56 to Page 7, Line 5: The following is not clear to me: '...the GPs were able to make a decision on suitable location of treatment in 88% of cases (n=1258 responses).' Does this mean that the GPs provided a 'Yes' or 'No' response (rather than 'Unsure') in 88% of responses, which equals 1,107 (0.88*1258) responses? Perhaps the sentence could be more clearly written.

We have clarified the numerator and denominator for the % calculation and we feel that the sentence reads clearly, in the context of the paragraph.

General: There are still instances in the text where suitability for 'treatment' rather than suitability for 'management' are referred to (e.g. Page 2, Line 26; Page 2, Line 31). The phrase 'could be seen' is also still used on occasion (e.g. Page 3, Line 22; Page 3, Line 27; Page 8, Line 40). The authors, in their response, indicated a preference for using the term 'management', so they may wish to ensure its consistent usage.

We thank the reviewer for pointing this out and we have altered these words accordingly.

REVIEWER 2

This paper now reads very well and I recommend acceptance for publication without further review, after the authors have considered the minor points below:

We thank this reviewer for these comments.

Abstract results: the main result is described as 'mean GP rating of appropriateness for primary care treatment was 43%', which I found slightly difficult to interpret, as it implies that the GP rating was a % score rather than a yes/no decision. How about: 'mean percentage of patients that GPs considered suitable for primary care treatment was 43%' instead?

We have altered the sentence as described, but have used the word ‘management’ to keep our terminology consistent, as recommended by reviewer 1.

The kappa agreement is given as 0.54, which is the result for the first pair of GPs. The result of 0.47 for the second pair should also be given here.

This has been added.

Abstract conclusion: the stronger agreement for paeds is discussed here, and so should probably be mentioned in the abstract results as well. Brief discussion of the low agreement between GPs & the implication that accurate assessment is difficult for individual patients would be useful here, space allowing.

This finding is already in the abstract, but we have changed the word ‘children’ to ‘paediatric’ to make this finding more visible to the general medical reader. Due to space constraints we have left the other aspects of the discussion in the paper rather than putting them in the abstract.

Methods para 1: '2 GPs demonstrated that this information was sufficient' would be more accurate as: '2 GPs considered that this information was sufficient'

We have altered this sentence as suggested.

Methods para 2: the new sentence starting 'We extracted data established as satisfactory....' took me a few readings to understand, and the authors may wish to clarify.

We have altered this sentence to clarify meaning. It now reads “We extracted data that had been established as satisfactory in the pilot for GP decision making from the record documented by a triage nurse, and transferred this to an electronic questionnaire”

Results para 1: clarify whether 'minors' area of ED' refers to minor injuries or children's area.

This is minor injuries. This has been clarified in the text.

Results para 2 & 4: see comment on abstract above about phrasing of main result.

We have altered these sentences as suggested from the comments above.

Discussion main findings para 1: The authors may wish to rephrase the sentence starting: 'differences between primary care...' as I found it difficult to interpret.

This sentence now reads “Differences between cases considered appropriate for primary care compared with those appropriate for the emergency department were most pronounced over the need for specialist review or admission, rather than need for investigations”

See comment under abstract conclusion above, that discussion of the low agreement between GPs & the implication that accurate assessment is difficult for individual patients should be added here. It is an important point for the future design of interventions in this area, that assessment of patient suitability is a complex task, and agreement even between experienced GPs is not high.